

Binäre Regression

Aufgabe 1: Wilkinson-Rogers-Notation

Seien A, B und C kategoriale Merkmale

- (i) Stellen Sie den folgenden Ausdruck in Wilkinson-Rogers-Notation dar

$$\begin{aligned} E(y|A, B, C) = & x_{A(1)}\beta_{A(1)} + \cdots + x_{A(I-1)}\beta_{A(I-1)} \\ & + x_{A(1)}x_{B(1)}\beta_{AB(1,1)} + \cdots + x_{A(I-1)}x_{B(J-1)}\beta_{AB(I-1,J-1)} \\ & + x_{B(1)}x_{C(1)}\beta_{BC(1,1)} + \cdots + x_{B(J-1)}x_{C(K-1)}\beta_{BC(J-1,K-1)} \end{aligned}$$

- (ii) Vereinfachen Sie den folgenden in Wilkinson-Rogers-Notation gegebenen Ausdruck:

$$(A + B + C) - (A * B - A.B)$$

Aufgabe 2: Binäre Modelle

Installieren Sie zunächst das Paket AER. (Hinweis: `?install.packages`).

Öffnen Sie anschließend den in AER zur Verfügung gestellten Datensatz `SwissLabor` durch den Befehl
`> data("SwissLabor", package="AER")`

Der Datensatz umfasst die Daten von 872 in der Schweiz lebenden Frauen und enthält folgende Variablen:

participation: berufstätig ja/nein (abh. Variable)
income: (log.) Einkommen unter Ausschluss beruflicher Verdienste
education: offizielle Ausbildungsdauer in Jahren
age: Alter in Jahrzehnten
youngkids/oldkids: Anzahl jüngerer/älterer Kinder
foreign: Schweizer Staatsbürgerin ja/nein

- (i) Plotten Sie zunächst die Variable **participation** auf geeignete Weise gegen **age**. Was fällt auf?
Hat das Alter einen signifikanten Einfluss auf die Berufstätigkeit (logistisches Modell)?
- (ii) Fitten Sie ein Logit-, ein Probit- und ein cLogLog- Modell für **participation** unter Berücksichtigung aller Einflussgrößen (ohne Interaktionen). Verwenden Sie, falls nötig, die R-Hilfe.
- (iii) Fitten Sie nun ein Logit-Modell, das neben allen Haupteffekten auch das Quadrat der Altersvariable beinhaltet. Wie beurteilen Sie dieses Modell im Vergleich zu den Modellen aus (ii)?